Audibility of spectral differences in head-related transfer functions

Hoffmann, Pablo F.F.; Møller, Henrik

Published in:
Proceedings of 120th AES Convention

Publication date:
2006

Audio Engineering Society
Convention Paper 6652
Presented at the 120th Convention
2006 May 20–23 Paris, France

Audibility of Spectral Differences in Head-Related Transfer Functions

Pablo Faundez Hoffmann[1], and Henrik Møller[1]

[1] *Department of Acoustics, Aalborg University, Aalborg, 9220-DK, Denmark*

Correspondence should be addressed to Pablo Faundez Hoffmann (pfh@acoustics.aau.dk)

ABSTRACT

The spatial resolution at which head-related transfer functions (HRTFs) are available is an important aspect in the implementation of three-dimensional sound. Specifically, synthesis of moving sound requires that HRTFs are sufficiently close so the simulated sound is perceived as moving smoothly. How close they must be, depends directly on how much the characteristics of neighboring HRTFs differ, and most important, when these differences become audible. Differences between HRTFs exist in the interaural delay (ITD) and in the spectral characteristics, i.e. the magnitude spectrum of the HRTFs. The present study investigates the audibility of spectral characteristics. To this purpose, binaural audibility thresholds of differences between minimum-phase representations of HRTFs are measured and evaluated.

1. INTRODUCTION

The sound field generated by a moving source can be synthesized at the ears of a listener as a sequence of stationary sound images that are changed over time. A single stationary sound is synthesized by convoluting an anechoic sound source with a head-related transfer function (HRTF). The HRTF is then updated at regular times, and a new convolution is computed in order to generate the next stationary sound. An important aspect of this type of synthesis, is the spatial resolution at which the stationary sound images are synthesized. The spatial resolution is usually determined by the spatial density of the measured HRTFs used to perform the synthesis. Assuming that widely spaced HRTFs have large differences in their characteristics, a change between sounds synthesized by them would most likely produce audible broken steps in the trajectory of the moving sound. It is not new that this problem might be alleviated by using interpolation algorithms [1], or a more dense set of HRTFs, but at the extra cost of computational power and memory requirements [2]. Therefore, it is of interest to investigate what is the largest angular separation between HRTFs that would allow a change between them without causing audible discrete steps. In order to address this question it seems appropriate to start by looking at the spatial resolution of the human auditory system.

The ability of human listeners to discriminate differences between sound sources located at different spatial positions, is commonly studied with emphasis on the detection of changes in the direction of a sound. In the literature, the threshold for discriminating the locations of stationary sounds is defined as the minimum audible angle (MAA) [3]. MAA thresholds for different directions have been obtained for real sound sources [3, 4], and also for virtual sound sources [5, 6, 7]. In the particular case of virtual sound, the MAA is of interest since it gives perceptual basis to define the necessary spatial resolution for the implementation of virtual auditory environments. However, since this threshold is based on localization judgments, it may also be of relevance to quantify how well humans discriminate differences between sound directions by means of any possible cue, and not necessarily those leading to a perceived change in direction. Since HRTFs are the functions used to control the directional information of a sound, discrimination of changes in the direction of a sound can be addressed by measuring the audibility of differences between the HRTFs used.

It is generally accepted that HRTFs can be approximated by a pure delay in cascade with a minimum-phase filter [2, 8]. The delay controls the inter-aural time difference (ITD), and the minimum-phase filter controls the magnitude of the HRTF's frequency response. In this manner, the influence of temporal and spectral characteristics can be studied separately. In previous studies, audibility thresholds for direct switching of temporal and spectral characteristics in HRTFs were measured [9, 10]. Generally, it was observed that time switching is more critical than spectral switching, meaning that listeners tend to be more sensitive to artifacts created by dynamically varying delays. Furthermore, from the experiment on spectral switching it was not completely clear that the estimated thresholds were solely due to the artifact created by the switching, or due to differences in the spectral content of the switched filters.

The aim of the present study is to estimate the angular separation at which differences between the spectral magnitude of the HRTFs become audible, while disregarding switching artifacts.

2. EXPERIMENTAL METHOD

2.1. Subjects

Four paid subjects participated in the listening experiment, one female and three males. Their ages ranged from 23 to 28. All subjects had previous experience on listening tests. Subjects' normal hearing was assessed by an audiometry, screening at less than 10 dB HL for frequencies ranging from 250 Hz to 4 kHz in octave steps, and less than 15 dB HL for 8 kHz.

2.2. Stimuli

Broadband pink noise (20 Hz - 16 kHz) generated at a sampling rate of 48 kHz was used as the basic signal. To simulate directional sound, HRTFs measured with a high resolution on an artificial head were used [11]. Eight directions were selected in the left half of the upper hemisphere. These directions are referred to as the *nominal directions*. Directions are given as (azimuth ϕ, elevation θ) in a polar coordinate system with horizontal axis and left-right poles (also referred to as the interaural-polar coordinate system). 90° and -90° azimuth correspond to left and right sides, 0° elevation to the frontal part of the horizontal plane, 180° elevation to the rear part of the horizontal plane, and 90° elevation to the upper part of the frontal plane. Four directions were selected in the median plane (0° azimuth; 0°, 44°, 136° and 180° elevation). Three directions were selected on a cone of confusion ((58°, 0°), (46°, 90°) and (54°, 180°)). These three directions were chosen to have the same ITD rather than being on the same geometrical cone, thus, their azimuth varies with elevation. The left direction, corresponding to (90°, 0°), was also selected. Fig. 1 shows the directions used in this experiment.

2.2.1. HRTF processing

Measured HRTFs were represented as minimum-phase FIR filters with the ITD calculated separately and inserted to the contralateral impulse response.

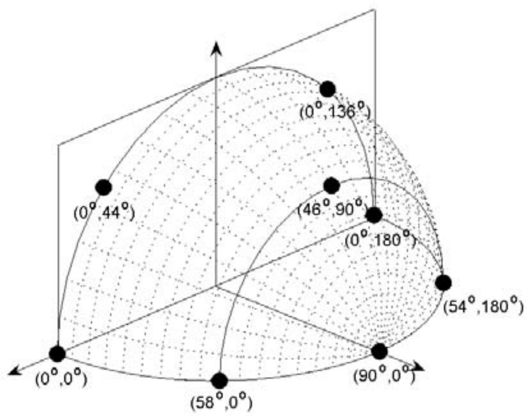

Fig. 1: Nominal directions selected at the left half of the upper hemisphere. Directions are specified by using an interaural-polar coordinate system.

Minimum-phase filters were set to a length of 1.5 ms (72 coefficients at 48 kHz). This length is sufficient to avoid audible effects of the truncation while keeping the appropriate spatial quality [12]. The DC value of each HRTF was set to unity gain (section 5.2 in [13]). ITD values were derived from the interaural differences in group delay of the excess-phase components of the HRTFs evaluated at 0 Hz [14]. The obtained ITDs were rounded to the nearest sample. Table 1 shows the values of the calculated ITD for the selected directions.

2.2.2. Playback System

Stimuli were played back using a PC equipped with a professional audio card RME DIGI96/8 PST. The digital output of the audio card was connected to a D/A converter with 16 bit resolution at a sampling rate of 48 kHz. From the D/A converter the signal went to a stereo amplifier (Pioneer A-616) modified to have a calibrated gain of 0 dB. A passive attenuator of -20 dB was then connected to the output of the amplifier in order to reduce its noise floor. Finally, the stereo output signal from the attenuator was delivered to the listener through a pair of Beyerdynamic DT-990 circumaural headphones.

Table 1: ITD values and corresponding nominal directions.

ITD (μs)	Nominal direction (ϕ, θ)
0	(0°,0°)
	(0°,44°)
	(0°,136°)
	(0°,180°)
-437.5	(58°,0°)
	(46°,90°)
	(54°,180°)
-625	(90°,0°)

2.2.3. Headphone Equalization

Two minimum-phase filters were employed in order to compensate for the left and right headphone transfer functions respectively. The design of the equalization filters was based on headphone transfer functions (PTFs) measured on blocked ear canal from 23 subjects. Five PTFs were obtained from each ear and subject; subjects were asked to reposition the headphones between measurements. PTFs were measured by using a maximum-length sequence technique (MLS) [15]. PTFs were averaged on a sound power basis, and a minimum-phase representation of the inverse of the average PTF was computed for each ear. More details on the measurement and equalization techniques can be found in [16].

By filtering the pink noise with the headphone equalization filters and setting fade-in and -out ramps of 10 ms each, the sound stimulus was ready to be delivered to the headphones. The overall gain of the system was calibrated so as the unprocessed pink noise simulated a free-field sound pressure level of 70 dB approximately.

2.3. Psychometric Method

A set of neighboring HRTFs was assigned to each nominal direction. These neighboring HRTFs were separated from the nominal direction at different angles. The selected angular separations were ±1°, ±4°, ±8°, ±12° and ±16° for all nominal directions except for the one directly in front of the listener. For this particular direction the selected angular

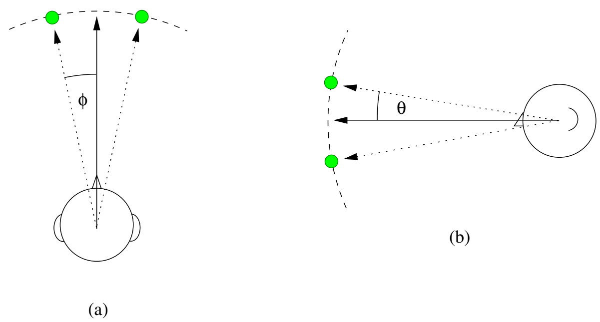

Fig. 2: Scheme of the two directional modes used in the experiment. The solid arrow represent the nominal direction. One pair of neighbor HRTFs, indicated by the dotted arrows, describes an arc using; (a) the azimuth mode with an angular separation of $\pm\phi$, (b) the elevation mode with an angular separation of $\pm\theta$.

separations were ±1°, ±2°, ±4°, ±6°, ±10°. The selection of these values was based on observations from pilot tests. Angular separations were extended on two directional modes. For one mode, angular separations were extended along the azimuth angle; for the other mode, angular separations were extended along the elevation angle. In the remaining part of this article they are referred to as the *azimuth mode* and the *elevation mode* respectively. Fig. 2 shows a graphical representation of these two directional modes for the nominal direction (0°, 0°). When the azimuth mode is used an arc is described along the horizontal plane. In case of employing the elevation mode, HRTFs along the median plane are used to describe the arc. Note that the midpoint of the arcs always corresponds to the nominal direction. For the particular case of nominal direction (90°, 0°), where the elevation mode cannot be applied, two azimuth modes were implemented. One mode in the horizontal plane spanning the angle horizontally, and the other mode in the frontal plane spanning the angle vertically.

A three-interval, two-alternative forced-choice method was employed (3I 2AFC). The duration of the stimulus (processed pink noise) was 300 ms and the inter-stimulus interval was also 300 ms. The stimulus presented on either the second or third interval differed from the other two. Subjects were asked to identify the interval that contained the different stimulus. They had to push one of two buttons to indicate a response. A feedback light was used to immediately show the correct response to the subjects. After a silence interval of 2 s a new trial was presented.

During a block of trials nominal direction and directional mode were held constant. For a single trial, an angular separation was selected and the minimum-phase HRTF of one end of the arc was used to filter two of the three sound intervals (one corresponding to the first interval). This filter was considered as the reference HRTF. The remaining sound interval was filtered with the minimum-phase HRTF corresponding to the other end of the arc. Note that the ITD associated to the nominal direction was used for both the reference and the different HRTFs. In case of using the azimuth mode, the spatial configuration of the reference-different HRTFs was randomly selected to be either left-right or right-left from the nominal direction. Similarly, for the elevation mode the reference-different HRTFs could be either above-below or below-above the nominal direction.

2.4. Experimental Design

Subjects were in a sound-insulated cabin with absorbing walls specially designed for psychoacoustical experiments. Once in the cabin subjects were provided with written instructions about the task they were to accomplish. They were then presented with a few trials in order to acquaint them with the task and the procedure. Posteriorly, they were presented with a set of sixteen trials for the training session. Since subjects had recently participated on similar listening tests and their performance was observed to be stable, no further training was considered necessary.

For the main experiment, all angular separations were repeated 15 times within a block of trials. The order in which angular separations were presented was randomized. Each condition corresponding to nominal direction - directional mode - angular separation was presented 30 times throughout the whole experiment. Thus, two blocks were necessary to complete data collection for one condition. A total of 2400 responses were obtained per subject (8 nominal directions x 2 directional modes x 5 angular separations x 30 repetitions). Data were collected during three sessions that were held on different days. Each session consisted of 12 blocks which were appropriately distributed so that one session lasted from about one hour and half to two hours.

3. RESULTS

Psychometric functions are shown for each subject in Fig. 3. The abscissa specifies the angular separation in degrees ($\pm°$), and the ordinate specifies the proportion of correct responses, giving an estimate of subjects' performance per condition. Each column represents results from one subject. Nominal directions are arranged in rows and grouped by corresponding ITD.

4. DISCUSSION

Audibility of spectral differences in HRTFs was estimated by measuring how well subjects could discriminate between minimum-phase HRTFs from different directions as a function of angular separation. It was assumed that differences between minimum-phase HRTFs would only yield variations in the spectral magnitude of the signal reaching the subjects' ears. It was expected that these variations provided subjects with the cues for performing the discrimination task. Note that in the strict sense one might argue that a change in the spectral phase would also occur. Fortunately, this aspect is of little concern since there is substantial evidence that the phase distortion introduced by minimum-phase representations of HRTFs is below the audible threshold [17, 18].

Results are generally consistent across subjects. Differences between minimum-phase HRTFs appear to systematically become more audible as the angular separation between them increases. This is observed from the fact that, for almost all subjects, performance tends to improve as the angular separation is incremented. For some subjects a decline in performance for larger angular separations was observed in some of the conditions. In this context it seems important to mention that variations in the spectral magnitude between HRTFs are rather complex across frequencies. In fact, for some of the HRTFs used in this experiment, e.g. neighboring HRTFs to the nominal direction (90°, 0°), it was observed that as the angular separation increased, differences between their spectral magnitudes did not increase but actually decreased for some frequency regions. If then the angular separation increased even more, the spectral differences at these frequency regions started to increase, thus, following a non-monotonic pattern. This situation was more evident for the contralateral component of the HRTFs. However, it is believed that the individual decays in performance at large angular separations might have most likely occurred due to attentional lapses. This is supported by the observed overlapping between the confidence intervals.

Observations from Fig. 3 show that for the nominal directions in the median plane (zero ITD), discrimination performance tended to be better for the elevation mode than for the azimuth mode. This suggests that, in the absence of temporal information, subjects seemed to be more sensitive to differences between minimum-phase HRTFs spanned along the elevation angle than those spanned along the azimuth angle. Note that in this particular experiment a change along the elevation angle corresponds to a physically correct situation, whereas a change along the azimuth is unnatural since the change in the corresponding interaural temporal information is not presented accordingly. Differences in performance between the two directional modes are more prominent for directions in the rear hemisphere than directions in the frontal hemisphere.

Discrimination performance among directions within the "cone of confusion" (ITD of -437.5 μs) showed no dramatic changes when using an elevation mode. For the condition corresponding to azimuth mode and nominal direction (46°, 90°), it was observed that subjects were less sensitive to differences than for directions to the front and the rear of the cone. These data are in agreement with the general knowledge that HRTFs actually change more smoothly for regions located above

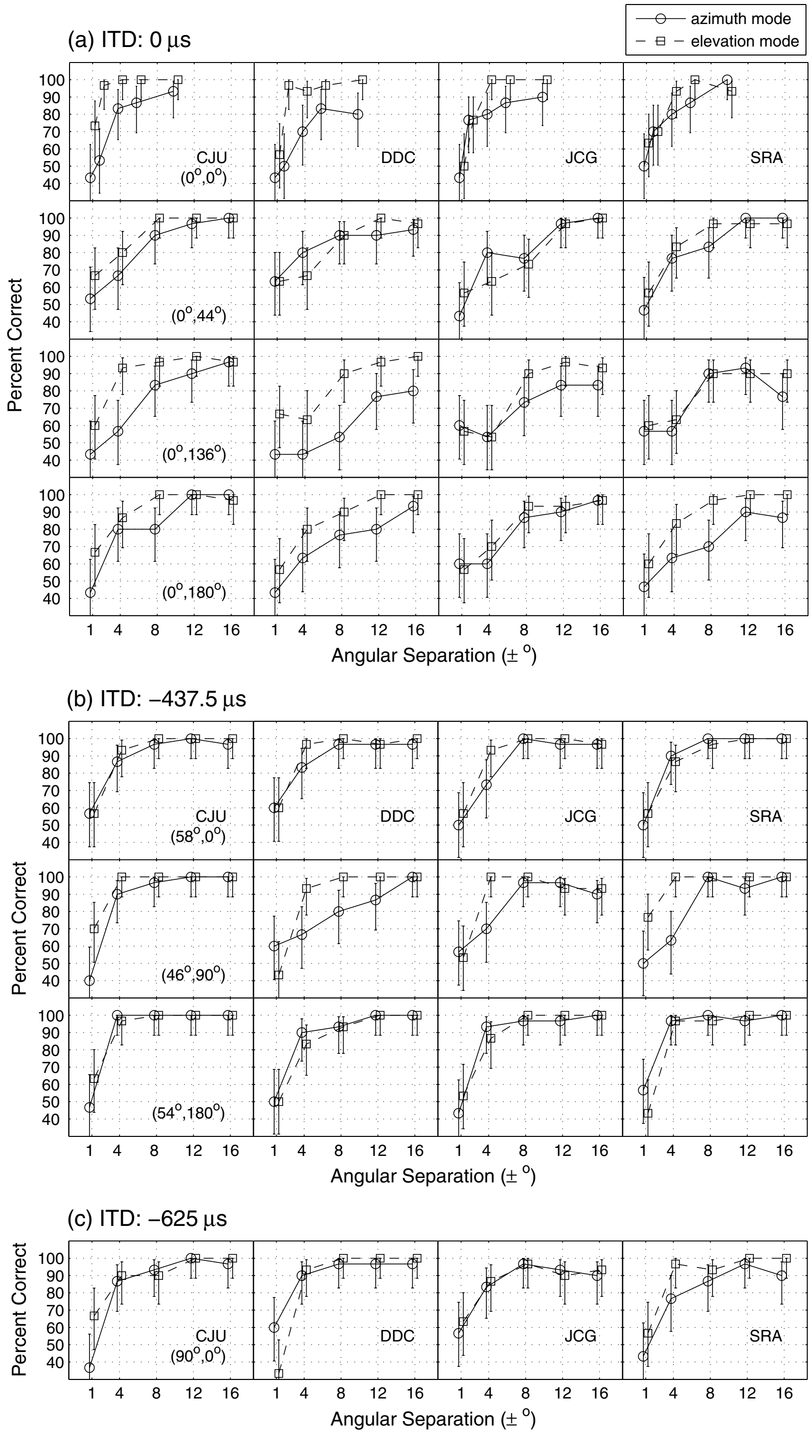

Fig. 3: Individual psychometric functions. Results from each subject are arranged column-wise, and nominal directions are organized in rows and grouped by common ITD. Two functions are presented representing performance for azimuth mode and elevation mode. Each point is calculated from 30 observations. Error bars represent 95% confidence intervals. Data have been slightly jittered for visualization purposes.

the head than regions near the horizontal plane or below. For the left nominal direction (ITD of -625 μs) performance was relatively equivalent between angular modes and among subjects.

In terms of differences between azimuth mode and elevation mode, results suggest that synthesis of moving sound along trajectories that incorporate changes in the elevation angle only, need a higher resolution than trajectories where azimuthal changes occur, i.e., trajectories where the ITD value also need to be updated.

Further experiments are currently work in progress in order to measure audibility of spectral differences between HRTFs for monaural hearing. In addition, experiments in which the spectral magnitude remains unchanged while only varying the interaural temporal information are also planned. This approach will complete the series of psychoacoustical experiments designed in order to evaluate audibility of differences in HRTFs when using a generic dataset. The expected benefit from these experiments is to obtain a perceptual baseline for making more informed decisions regarding engineering compromises in the implementation of dynamic three-dimensional sound.

5. ACKNOWLEDGMENTS

Economic support from the Danish Technical Research Council and the Research Council for Technology and Production Science is greatly acknowledged.

6. REFERENCES

[1] K. Hartung and J. Braasch. Comparison of different methods for the interpolation of head-related transfer functions. In *Proceedings of the 16th Audio Engineering Society (AES International Conference on Spatial Sound Reproduction*, Rovaniemi, Finland, 1999.

[2] Jean-Marc Jot, V. Larcher, and O. Warusfel. Digital signal processing in the context of binaural and transaural stereophony. In *98th AES Convention*, Paris, France, February 25-28 1995. Preprint 3980.

[3] A. W. Mills. On the minimum audible angle. *J. Acoust. Soc. Am.*, 30(4):237–246, December 1958.

[4] David R. Perrott and Kourosh Saberi. Minimum audible angle thresholds for sources varying in elevation and azimuth. *J. Acoust. Soc. Am.*, 87(4):1728–1731, April 1990.

[5] R. L. McKinley, M. A. Ericson, David R. Perrott, Robert H. Gilkey, Douglas Brungart, and Frederic L. Wightman. Minimum audible angles for synthesized location cues presented over headphones. In *J. Acoust. Soc. Am.*, volume 92, page 2297. October 1992.

[6] R. L. McKinley and M. A. Ericson. *Binaural and Spatial Hearing in Real and Virtual Environments*, chapter in R. H. Gilkey, T. R. Anderson (edt.) : Flight Demonstration of a 3-D Auditory Display, pages 683–699. Lawrence Erlbaum Associates, 1997. ISBN: 0-8058-1654-2.

[7] Robert S. Bolia and Alan. D. Musicant. Monaural and binaural minimum audible angles for virtual sound sources. In *J. Acoust. Soc. Am.*, volume 105, pages 1024–1025.

[8] A. Kulkarni, S. K. Isabelle, and H. S. Colburn. On the minimum-phase approximation of head-related transfer functions. In *Applications of Signal Processing to Audio and Acoustics*, pages 84–87. IEEE ASSP Workshop, 15-18 Oct 1995.

[9] P. F. Hoffmann and H. Møller. Audibility of time switching in dynamic binaural synthesis. In *118th AES Convention*, Barcelona, Spain, May 27-31 2005. Preprint 6326.

[10] P. F. Hoffmann and H. Møller. Audibility of spectral switching in head-related transfer functions. In *119th AES Convention*, New York, USA, October 7-10 2005. Preprint 6537.

[11] B. P. Bovbjerg, F. Christensen, P. Minnaar, and X. Chen. Measuring the head-related transfer functions of an artificial head with a high directional resolution. In *109th AES Convention*, Los Angeles, California, USA, September 22-25 2000. Preprint 5264.

[12] J. Sandvad and D. Hammershøi. What is the most efficient way of representing HTF filters? In *Proceedings of Nordic Signal Processing Symposium*, NORSIG '94, Lesund, Norway, June 2-4 1994.

[13] D. Hammershøi and H. Møller. *Communication Acoustics*, chapter in Jens Blauert (edt.) : Binaural Technique, Basic Methods for Recording, Synthesis and Reproduction, pages 223–254. 2005. Springer Verlag, ISBN 3-540-22162-x.

[14] P. Minnaar, J. Plogsties, S. K. Olesen, F. Christensen, and H. Møller. The interaural time difference in binaural synthesis. In *108th AES Convention*, Paris, France, February 19-22 2000. Preprint 5133.

[15] D. D. Rife and J. Vanderkooy. Transfer-function measurement with maximum-length sequence. *J. Audio Eng. Soc.*, 37(6):419–444, 1989.

[16] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen. Transfer characteristics of headphones measured on human ears. *J. Audio Eng. Soc.*, 43(4), April 1995.

[17] A. Kulkarni, S. K. Isabelle, and H. S. Colburn. Sensitivity of human subjects to head-related transfer function phase spectra. *J. Acoust. Soc. Am.*, 105(5):2821–2840, May 1999.

[18] P. Minnaar, H. Møller, J. Plogsties, S. K. Olesen, and F. Christensen. The audibility of all-pass components in binaural synthesis. *J. Audio Eng. Soc.* In Preparation.